المجلة العراقية للعلوم الإحصائية

www.stats.mosuljournals.com

تحسين تنبؤ التعلم الآلي باستخدام خوارزمية الفراولة

نورا سهيل اسماعيل حاوا و صالح مؤيد شاكر جرجيس

قسم الاحصاء والمعلوماتية، كلية علوم الحاسوب والرياضيات، جامعة الموصل ، الموصل ، العراق

معلومات النشر

تاريخ المقالة:
تم استلامه في 20 ايلول 2021
تم القبول في 20 تشرين الاول 2021
متاح على الإنترنت في 1 حزيران 2022

الكلمات الدالة:

انحدار المتجه الداعم، التعليم الالي ، الخوارزميات، خوارزمية الفراولة.

المراسلة:
نورا سهيل اسماعيل حاوا
noora.csp148@student.uomosul.edu.iq

الخلاصة

في هذا البحث تم استخدام نموذج انحدار المتجه الداعم ((Support Vector Regression(SVR) حيث يعرف بانه خوارزمية او نموذج خطي يستخدم للتنبؤ بنموذج معين. يعتمد كفاءة اداء أسلوب SVR على اختيار معلماتهِ الفائقة. تم في هذا البحث استخدام أسلوب SVR مع خوارزمية الفراولة (Strawberry Algorithm) *وهي الخوارزمية المقترحة للحصول على أفضل توليفة للمعلمات الفائقة*. وتم استخدام معيار جذر متوسط مربعات الخطأ Root Mean Squares Error (RMSE) لمقارنة النتائج التي تم الحصول عليها باستخدام الخوارزمية المقترحة مع بعض الخوارزميات الشائعة وهي خوارزمية البحث الشبكة (Grid Search)، الخوارزمية الجينية (Genetic Algorithm)، خوارزمية سرب الطيور (Particle swarm optimization)، وخوارزمية التلدين (Simulated Annealing algorithm) . كما تم استخدام ثلاثة طرائق للاختيار في خوارزمية الفراولة وهي طريقة اختيار عجلة الروليت(The roulette wheel) اختيار النخبة (Elite)واختيار عجلة الروليت مع النخبة معاً (Roulette & Elite) . وتم اختبار اداء الخوارزمية من خلال بيانات تجريبية وحقيقية. وقد أظهرت النتائج تفوق خوارزمية الفراولة على الخوارزميات الشائعة في اختيار أفضل توليفة للمعلمات الفائقة. كما أظهرت النتائج ان طريقة اختيار عجلة الروليت (the roulette wheel) أفضل الطرائق التي اعطت نتائج جيدة مقارنة بالطرائق الأخرى في الجانبين التجريبي والتطبيقي.

DOI: 10.33899/IQJOSS.2022.174327 , ©Authors, 2022, College of Computer Science and Mathematics, University of Mosul.
This is an open access article under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

مقدمة Introduction

يعتبر الانحدار(Regression) فرع من فروع علم الإحصاء، وهو اداة احصائية الاكثر استخداما من حيث بناء نموذج احصائي يعمل على تقدير العلاقة بين متغيرين وهما المتغير التابع (اي متغير الاستجابة) مع متغير واحد او عدة متغيرات توضيحية **(Explanatory variables)**. اذ يقوم بدراسة العلاقة بين المتغيرين لغرض التقدير والتنبؤ، ويستخدم الانحدار في عدة مجالات وفي مختلف العلوم والتطبيقات. مؤخرا لجأ الباحثون الى استخدام التعليم الالي الذي يعتبر فرع من فروع الذكاء الاصطناعي (Artificial Intelligence) والذي يهتم بدراسة التطبيقات البرمجية للحصول على نتائج اكثر دقة والذي يهتم بتصميم وتطوير الخوارزميات باستخدام التحليل الإحصائي (Analysis statistical) .يعرف انحدار المتجه الداعم Support Vector Regression (SVR) بانه تعليم آلي او خوارزمية او نموذج آلي الذي يبني على تنبؤ نموذج خطي عن طريق التصغير. ويكون مبدا عمل SVR من خلال تقسيم البيانات الى مجموعتين، المجموعة الاولى تسمى بيانات التدريب والمجموعة الثانية تسمى بيانات الاختبار وتعرف ايضا ببيانات النموذج او بيانات الهدف والتي من خلالها يمكن قياس دقة التنبؤ. هنا سوف نقترح دمج اسلوب SVR مع احدى الخوارزميات المستوحاة من الطبيعة وهي الخوارزمية المقترحة، خوارزمية نبات الفراولة Strawberry Algorithm (SBA)للحصول على أفضل توليفة للمعلمات الفائقة من خلال الحصول على اقل جذر لمتوسط مربعات الخطأ(RMSE) Root Mean Squares Error . تم

تطبيق خوارزمية الفراولة والاساليب الثلاثة للاختيار للبيانات الاربعة التجريبية، والجانب التطبيقي ناقش تطبيق بيانات حقيقة والمتمثلة بأحد الامراض المزمنة وهو مرض ضغط الدم (Blood Pressure) حيث تم تطبيق الجانبين باستخدام برنامج R الاحصائي.

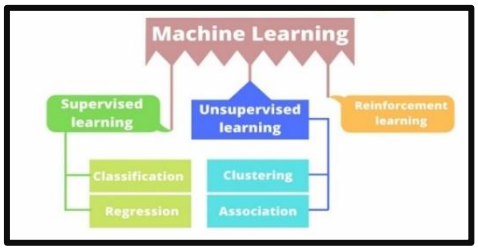
1. هدف البحث: Objective of Research

تهدف الدراسة الى الحصول على أفضل توليفة من المعلمات الفائقة لانحدار المتجه الداعم (SVR). حيث تم توظيف خوارزمية نبات الفراولة (SBA) والاعتماد على معيار جذر متوسط مربعات الخطأ RMSE مع التطبيق على بيانات تجريبية وبيانات حقيقة من اجل اثبات كفاءة هذه الخوارزمية للحصول على أفضل توليفة من المعلمات الفائقة.

2. التعليم الالي: Machine learning

يعتبر التعليم الالي فرع من فروع الذكاء الاصطناعي (Artificial Intelligence) الذي يهتم بدراسة التطبيقات البرمجية للحصول على نتائج اكثر دقة، وقد استخدم في تصميم وتطوير الخوارزميات للبيانات المدخلة وباستخدام التحليل الإحصائي (statistical Analysis) للتنبؤ بالمخرجات (Langley & Carbonell, 1984).

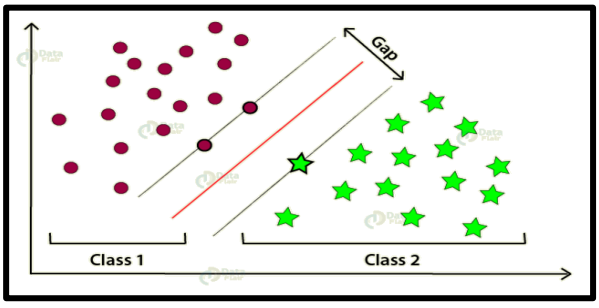
وكما يقسم التعليم الالكتروني (Machine learning) الى قسمين التعليم بأشرف(Supervised machine learning) والتعليم بدون اشراف (Unsupervised machine learning). والتعليم بأشرف سيقسم ايضا الى مجالين، الأول هو مجال التصنيف (classification) والثاني هو مجال الانحدار والتنبؤ (Regression) والذي سيكون في تخصص دراستنا، والشكل (1) يوضح تقسيم التعليم الالي(Anikesh, 2018) .



الشكل (1) يوضح تقسيم التعليم الالى

3. آلة المتجه الداعم Support Vector Machine

تم اكتشاف خوارزمية آلات المتجهات الداعمة SVM من قبل الباحثين فلاديمير فابنك وأليكسى شيرفونينكيز. وطورت من قبل العالم (Vapnik, 1995) ويعتمد في عملهِ على مجموعة من المتغيرات والعوامل التي تؤثر على الحل النهائي. فمثلا يعتمد على المعلمات الاساسية مثل المستوى الفاصل (Hyperplane) ومضاريب لانكراج (Lagrange multiplier) التي تؤثر على دقة الحل. (قاسم، 2009). والية عمل الة المتجه الداعم هي البحث على أحسن طريقة لفصل البيانات إلى مجموعتين من خلال المستوى الفائق (hyperplane) الذي يفصل بينهما (Burges, 1998).



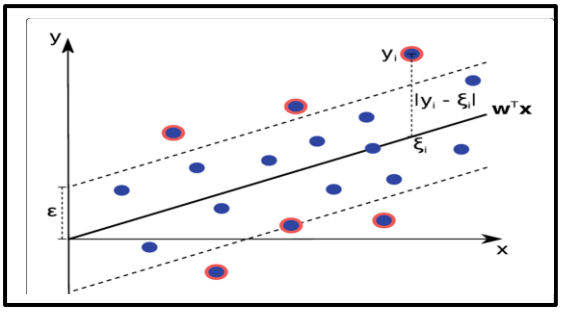
الشكل (2) يوضح الية عمل **آلة المتجه الداعم SVM**

4. انحدار المتجه الداعم Support Vector Regression (SVR)

يعد انحدار المتجه الداعم (SVR) بمثابة تعليم آلي او خوارزمية او نموذج آلي الذي يبنى على تنبؤ نموذج خطي عن طريق التصغير. حيث يستخدم جزء من البيانات والتي تُسمى بمجموعة بيانات التدريب. كما يقوم SVR ببناء نموذجًا خطيًا للتنبؤ بـ(y)، بعد رسم بيانات التدريب على مساحة الميزة ذات بعد عالي

$$f(w,x) = w^{*}\varphi(x) + b \qquad (1)$$

حيث ان w هو متجه معاملات الانحدار، $\varphi(x)$ هو رسم او مخطط في فضاء عالي الابعاد، و b هو حد التحيز، ويتم حساب التنبؤ بالخطأ باستخدام دالة الخسارة الحساسة $f(x)$ (insensitive loss function) (Rosenbaum et al., 2013).والشكل (2-5) يوضح خوارزمية انحدار المتجه الداعم (SVR) (Jacobs, 2012).



الشكل (3) يوضح خوارزمية انحدار المتجه الداعم (SVR)

والمعادلة ادناه توضح دالة الخسارة (loss function):

$$L(y, f(w,x)) = \begin{cases} 0 & if\ |y_i - f(w,x)| \le \varepsilon \\ |y_i - f(w,x)| - \varepsilon & otherwise \end{cases} \quad (2)$$

والهدف من استخدام SVR هو إيجاد $f(w,x)$ التي تتوافق مع انحراف ε لجميع بيانات التدريب من نموذج التنبؤ وفي نفس الوقت الذي يجعل $f(w,x)$ مسطح قدر الإمكان.

يمكن صياغة المشكلة كمشكلة تحسين محدبة:

$$\text{Minimize:} \quad \frac{1}{2}\|w\|^2 \quad (3)$$

$$\text{Subject to:} \begin{cases} y_i - f(x_i, w) - b \le \varepsilon + \delta_i^* \\ f(x_i, w) + b - y_i \le \varepsilon + \delta_i \end{cases} \quad (4)$$

في بعض الأحيان لا يمكن التنبؤ بكل بيانات التدريب داخل الانحراف ε. ولذلك، فإننا نقدم المتغيرات الرخوة (Slack Variables) لتحمل بعض التدريبات التي تقع خارج أنبوب حساس ε.

$$\text{Minimize: } = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\delta_i + \delta_i^*) \quad (5)$$

$$\text{Subject to:} \begin{cases} y_i - f(x_i, w) - b \le \varepsilon + \delta_i^* \\ f(x_i, w) + b - y_i \le \varepsilon + \delta_i \\ \delta_i^*, \delta_i \ge 0 \end{cases} \quad (6)$$

حيث C هي معلمة تنظيمية ثابتة تحدد المفاضلة بين تسطح $f(x,w)$ والخطر التجريبي. وان ξ_i, ξ_i^* هي متغيرات التي تقيس انحرافات أكبر من ε ،وان مشكلة التحسين تم تحويلها إلى مشكلة مزدوجة. ومن ثم الى استخدام معادلة Lagrange. ويمكن حل المشكلة وذلك من خلال التعبير عن الحل على النحو التالي:

$$f(x) = \sum_{i=1}^{n}(a_{i-} a_i^*)\varphi(x_i, x) + b \quad (7)$$

حيث a_i, a_i^* هي مضاعفات لاغرانج.

ويعتمد الحل النهائي فقط على حاصل الضرب وليس هنالك حاجة لتعيين البيانات بشكل صريح وليس من الضروري معرفة رسم الدالة.

وباستخدام دالة النواة (kernel) يمكننا استبدال حاصل الضرب بدالة لحساب حاصل الضرب لنقاط البيانات المرسومة في مساحة عالية الأبعاد. ليصبح الحل

$$f(x) = \sum_{(i=1)}^{n}(a_{i-} a_i^*)k(x_i, x) + b \quad (8)$$

حيث أن $k(x_i, x)$هي دالة kernel.

وباستخدام دالة kernel، يبقى حساب مشكلة التحسين دائمًا في مساحة الادخال بدلاً من مساحة التعيين لذلك يمكن أن يعمل أيضًا في مساحة تعيين ذات بُعد لا نهائي. وكما يمكن ايضا استخدام دوال kernel المختلفة مع SVM.(Engel et al., 2002). ويعتمد أداء (SVR) على الاختيار الصحيح للمعلمات الفائقة hyper parameters التي هي ε و c و σ.وأن المعلمة الفائقة ε تتحكم بعرض أنبوب (ε– insensitive) حول نموذج التنبؤ. (Yeh et al., 2011).

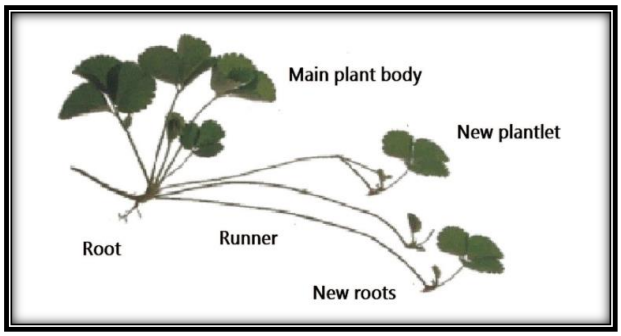
4. مفهوم الخوارزميات

الخوارزمية هي سلسلة من الخطوات الرياضية والمنطقية المتسلسلة لحل مشكلة معينة. وفي الاونة الاخيرة قام الباحثين باستخدام الخوارزميات المستوحاة من الطبيعة وتطويرها لحل وتحسين العديد من المشاكل الهندسية والرياضية المعقدة في ابحاثهم وللحصول على افضل حل (Merrikh–Bayat, 2014a).

وهنالك العديد من الخوارزميات المستوحاة من الطبيعة، ولكل خوارزمية مميزاتها وطريقة استخدامها ومن هذه الخوارزميات هي الخوارزمية الجينية **Genetic algorithm)**التي تعتبر من اول الخوارزميات التي ظهرت والتي اكتشفها الباحث Jon Holand عام 1975، ويكون اساس عمل الخوارزميات بتطبيق المتغيرات العشوائية (Random variables) والتعامل معها. (Aamir et al., 2016).

5. نبات الفراولة في الطبيعة

ينمو نبات الفراولة من خلال الجذور والسيقان، ويكون على شكل عقد وينمو في كل عقد جذر وساق ويسمى هذا نبات الام(Mother plant) ، والساق يكون زاحفا في الارض وينتج في محاور الأوراق وينمو من نبات الأم. ومن ثم يكون الثانية ليكون مصنع جديد يسمى نبات الابنة ومن ثم يظهر ساق جديد مرة ثانية لتوليد نبات ابنة جديدة وهكذا عند نمو النبات الام بشكل كامل يتم فصل النباتات الوليدة (الابنة) عن النبات الاصلي (الأم) حيث كل منها يواصل حياته بشكل فردي كنبات أم جديدة. حيث يمتلكان الام والابنة بنفس الصفات (Aamir et al., 2016). والشكل (4) يوضح نبات الفراولة (Konidala & Abstract., 2021).



الشكل (4) يوضح نبات الفراولة

6. مية الفراولة Strawberry algorithm

قام الباحثين بفهم العلاقة بين التعلم الآلي ونظام الطبيعة وإيجاد الطرق للحصول على أداء امثل عمل الباحثين بأسلوب الدمج بين الخوارزميات المستوحاة من الطبيعة مع التعلم الالي. (Merrikh–Bayat, 2014a).

يبحث نبات الفراولة عن موارد الطبيعة مثل المياه والمعادن من خلال تطوير السيقان والجذور والشعيرات الجذرية ويسمى هذا البحث بالبحث العالمي Globa) (search والبحث المحلي (Local search) على التوالي. وتعتبر السيقان والجذور أدوات البحث، ويقصد بالبحث العالمي وصول الساق الى المكان الابعد للبحث عن موارده الطبيعية اما البحث المحلي هو وصول الجذر الى المكان الاقرب الذي تتواجد به موارد المياه (Aamir et al., 2016).

7. خوارزمية الفراولة رياضياً

هناك عدة طرائق في علم الرياضيات لحل مشاكل التحسين لكن تكون مقعدة وتحتاج الى وقت للحصول على الحل الامثل. وتكون مشكلة التحسين من الطرق الاسهل، ويقصد بها الحصول على القيم الصحيحة لمجموعة من المتغيرات والعثور على أكبر او أصغر قيمة لمؤشر الأداء وكما تعتمد العديد من التطبيقات الهندسية على حل مشكلة التحسين.

تعمل خوارزمية الفراولة رياضيا لحل مشكلة التحسين الرقمي المقيد كالاتي:

$$\min f(x), \quad x_l \le x \le x_u \tag{9}$$

حيث ان $f: R^m \to R$ تمثل دالة هدف او التكلفة (cost) ذات أبعاد m، والتي يتم تصغيرها، وان $x \in R^m$ حيث ان X هو متجه الحل وان $x_l\ x_u \in R^m$ حيث ان $x_l\ x_u$ تعبر عن الحد الأعلى والأدنى لحدود المتغير. ويتم تنفيذ خوارزمية الفراولة لحل مشكلة التحسين من خلال تكوين N والتي تعرف باسم النبات الام (Mother plant) وبعدها يتم إنشاء نقطتين (واحدة قريبة جدًا من النبات الأم وهي الجذر (root) والأخرى بعيدة عنه تسمى الساق (runner).

وسوف تقوم الخوارزمية في إيجاد الحد الأدنى والاعلى. ويكون نبات الأم N في التكرار j ويمثل النقاط $x_{j(i)} \in R^m$ حيث $j = 1,2,3, \dots \dots . n$. يمكن تمثيل المصفوفة التي تتكون من السيقان والجذور التي ينتجها نبات الأم N من خلال الاتي:

$$x_{prop}(i) = [x_{root}(i)\ \ x_{runner}(i)] = [x(i)\ \ x(j)] + [d_{root} r1\ \ d_{run} r2] \tag{10}$$

حيث ان:

$$X_{(i)} = [x_1(i) x_2(i) \dots \dots . x_N(i)],$$

$$x_{prop}(i) = [x_{1prop}(i)\ x_{2prop} \dots \dots . x_{2Nprop}(i) \tag{11}$$

$$x_{root}(i) \in R^{m*N},\ x_{runner}(i) \in R^{m*N}$$

$$x_{root}(i) = \left[x_{1,root}(i) x_{2,root}(i), \dots \dots . . x_{N,root}(i)\right], \tag{12}$$

$$x_{runner}(i) = \left[x_{1,runer}(i) x_{2,runer}(i), \dots \dots . . x_{N,runer}(i)\right] \tag{13}$$

$r_1 r_2 \in R^m$ هي مصفوفات عشوائية مدخلاتها عبارة عن أرقام عشوائية مستقلة تتبع التوزيع المنتظم للفترة $[-0.5, 0.5]$ او توزيع اخر من التوزيعات المستمرة او المتقطعة. اما d_{root} و d_{runner} يمثلان المسافة لكل من الجذور والسيقان عن نبات الأم، اي العقد التي تتمثل بـ N، حيث ان الجذر الواحد يكون القريب والساق الواحد يكون الابعد عن نبات الام، ودائما تكون $d_{runner} \gg d_{root}$، وبعد حساب المصفوفة $X_{prop}(i)$ سوف نستخدم احد أساليب الاختيار والتي منها اما عجلة الروليت (Roulette wheel) او اختيار النخبة (elite-selection) (Merrikh-Bayat, 2014b).

10. **الخوارزمية الجينية:** Genetic algorithm

وهي خوارزمية مستوحاة من نظرية التطور الطبيعي لتشارلز داروين وتعتبر اسلوب من اساليب الذكاء الاصطناعي وقد قام بتطويرها العالم John Holland وزملائه في الستينيات والسبعينيات في الجامعة الامريكية في مشيكان عام 1975. ويعتبر John Holland أول من استخدم التقاطع وإعادة التركيب والطفرة والاختيار في دراسة الأنظمة التكيفية والاصطناعية والتي تستخدم لحل مشاكل التحسين (قاسم، 2018). وكما تحاكي الخوارزمية الجينية عملية الانتقاء الطبيعي حيث يتم اختيار الافراد الاكثر لياقة للتكاثر من أجل إنتاج نسل من الجيل التالي. (الكلاك & شعبان، 2008).

11. خوارزمية سرب الطيور: Particle Swarm Optimization Algorithm

تعد خوارزمية سرب الطيور (PSO) من الخوازميات المستوحاة من الطبيعة وطورت من قبل الباحثين إبرهارت وكينيدي في عام 1995، مستوحاة من السلوك الاجتماعي لتدفق الطيور أو تعليم الأسماك. وتمت تهيئة النظام بمجموعة من الحلول العشوائية وعمليات البحث عن الحل الامثل(Optima) عن طريق تحديث الأجيال. (Mu & Nandi, 2009).

12. خوارزمية التلدين المحاكاة Simulated annealing algorithm

تحاكي خوارزمية محاكاة التلدين عملية التبريد البطيء للمعدن المنصهر لتحقيق الحد الأدنى من قيمة الدالة في مشكلة التصغير. تتم محاكاة ظاهرة التبريد للمعدن المنصهر من خلال إدخال معامل شبيه بدرجة الحرارة والتحكم فيها باستخدام مفهوم توزيع احتمالية Boltzmann. يشير توزيع احتمالية Boltzmann إلى أن الطاقة (E) لنظام في التوازن الحراري عند درجة الحرارة T يتم توزيعها بشكل احتمالي (Thede, 2004). تعتمد طريقة محاكاة التلدين على محاكاة التلدين الحراري للمواد الصلبة التي تكون شديدة التسخين(Moriguchi et al., 2015).

13. خوارزمية الشبكة Grid search

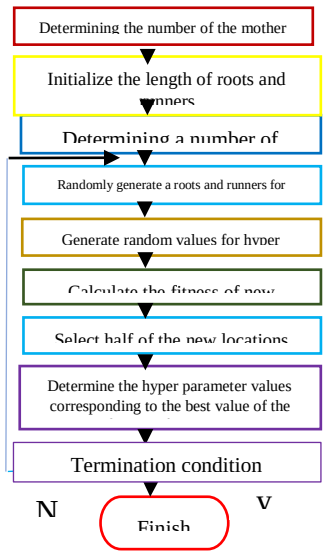
تعتبر من خوارزميات التحسين وتستخدم لضبط المعلمات الفائقة (hyper parameter) في نماذج التعلم الآلي والحصول على افضل المعلمات التي تجعل النموذج اكثر دقة (Kim, 1997).

واما بحث الشبكة (Grid search) هو عملية معالجة البيانات لتكوين المعلمات المثلى لنموذج معين اعتمادًا على نوع النموذج ويتطلب وجود معلمات معينة. (Evan Lutins, 2017).

14. الأسلوب المقترح في خوارزمية الفراولة

استخدمت خوارزمية الفراولة كأسلوب لاختيار أفضل توليفة من معلمات أسلوب SVR، حيث لم يتم استخدام هذه الخوارزمية من قبل في عملية اختيار المعلمات الفائقة التي تعود لأسلوب SVR. ان خوارزمية الفراولة وكما وضحها (Bayat,2014) تتميز بأسلوبها من حيث البحث المحلي والبحث العالمي للحصول على أفضل حل. وتتضمن هذه الخوارزمية على اثنان من مواقع الحل والمتمثل بالجذر والساق العائدين لنبات الام. وقد وضح (Bayat,2014) ان عملية اختيار الحل الأمثل تكون من خلال اختيار اما الساق او الجذر أي الحصول على حل واحد فقط كما في باقي الخوارزميات الأخرى. لذلك يستخدم أحد أساليب الاختيار والمتمثلة بأسلوب عجلة الروليت او أسلوب اختيار النخبة او الجمع بينهما كما فضل هذا الباحث. كما تم تطبيق التوصيات التي أوصى بها الباحث (Bayat,2014) من حيث استخدام عدة قيم للجذر وللساق، وعدة جذور وسيقان، إضافة الى ذلك تم استخدام أكثر من نبات ام واحد حيث تم استخدام N=1,2,3,4.ويمكن توضيح الية الربط بين خوارزمية الفراولة وانحدار المتجه الداعم من خلال المخطط الموضح في الشكل (5).

ان الأسلوب الذي تم اعتمادهُ في الية تطبيق أساليب الاختيار عند تطبيق أسلوب روليت ويل واسلوب النخبة معاً يختلف عما هو موجود في خوارزمية الفراولة، حيث ان في خوارزمية الفراولة يتم اختيار افضل دالة هدف من بين قيم دالة الهدف لكل من الجذر والساق، أي يتم اختيار قيمة دالة الهدف اما للجذر او للساق، اما هنا فقد تم تجزئة المجتمع الى نصفين، حيث تم الاختيار من النصف الأول باستخدام أسلوب عجلة الروليت، اما النصف الاخر فقد تم اختيار افضل قيمة لدالة الهدف باستخدام أسلوب النخبة، بذلك سيكون هناك قيمتين لدالة الهدف، عندئذٍ سيتم اختيار اقل قيمة منهما وتعتبر هي القيمة الأفضل. ونفس الأسلوب يطبق على قيم دالة الهدف التي تعود الى الساق. والهدف من هذا الأسلوب المقترح من قبل الباحث هو من اجل استخدام المعلمات نفسها التي أعطت اقل قيمة لدالة الهدف في التكرار التالي لكل من الجذر والساق. كما ان خوارزمية الفراولة تستغرق وقت أكثر من الخوارزميات الشائعة التي سوف نقارن بها مثل الخوارزمية الجينية، خوارزمية البحث الشبكي، خوارزمية التلدين المحاكاة، خوارزمية سرب الطيور. ويعود ذلك لسببين، الاول ان خوارزمية الفراولة تبحث في اتجاهين هما الجذور والسيقان عكس الخوارزميات الاخرى التي تبحث في اتجاه واحد. والسبب الثاني هو ان الجذور والسيقان التي تم استخدامهم كان 100 لكل منهما، اي 100 جذر و100 ساق هذا إضافة الى انهُ تم استخدام أكثر من عقدة N لكل 100 جذر وساق.



الشكل (5) مخطط لألية الربط بين خوارزمية الفراولة

الجانب التجريبي

15. المقدمة: Introduction

في الجانب التجريبي تم استخدام بيانات تجريبية من اجل اختبار كفاءة الأسلوب المقترح او الخوارزمية المقترحة حيث تم استخدام هذه البيانات من قبل اغلب الباحثين في بحوثهم العلمية (Rojas et al., 2017). وقد تم اخذ هذه البيانات من الموقع (UCI) وهي (Slump, Forest fires, Boston, Qsar_Aquatic_Toxicity). وتم تحليلها بالبرنامج الاحصائي R وتم انشاء برامج للخوارزميات المذكورة اعلاه وتطبيق على ما جاء في الجانب النظري من هذا

البحث وباستخدام الخوارزمية المقترحة (خوارزمية الفراولة) لأساليب الاختيار الثلاثة ومقارنتها مع الخوارزميات الشائعة الاخرى باستخدام معيار جذر توسط مربعات الخطأ (RMSE) للحصول على أفضل توليفة للمعلمات الفائقة لانحدار المتجه الداعم SVR. والجدول (1) ادناه يوضح الوصف لهذه البيانات التجريبية من حيث حجم العينة والمتغيرات التوضيحية.

الجدول (1) يوضح حجم العينة وعدد المتغيرات لكل مجموعة من البيانات التجريبية

اسم البيانات	Slump	Boston	Aquatic	Forestfire
حجم العينة	103	506	545	517
المتغيرات التوضيحية	8	14	9	11

ويمكن تمثيل عدد العقد المستخدمة وهي N1,N2,N4 وكل عقدة تحتوي على 100 جذر و 100 وساق حيث تم اخذ ثلاثة اطوال للجذور والسيقان، حيث كانت اطوال الجذور (0.6 ,0.4, 0.2) اما اطوال السيقان فكانت (4.5 ,3.5, 2.5) لحصول على نتائج مثلى. والجدول (2) يوضح ذلك.

الجدول (2) يوضح كل عقدة N مع اطوال الجذور والسيقان

Mother plane	length Roots and runners
N11	1= One Mother plant, 1= (0.2) root length & (2.5) runner length
N12	1= One Mother plant ,2= (0.4) root length & (3.5) runner length
N13	1= One Mother plant, 3= (0.6) root length & (4.5) runner length
N21	2= Two Mother plants, 1= (0.2) root length & (2.5) runner length
N22	2= Two Mother plants, 2= (0.4) root length & (3.5) runner length
N23	2= Two Mother plants, 3= (0.6) root length & (4.5) runner length
N41	4= four Mother plants, 1= (0.2) root length & (2.5) runner length
N42	4= four Mother plants, 2= (0.4) root length & (3.5) runner length
N43	4= four Mother plants, 3= (0.6) root length & (4.5) runner length

تم تطبيق خوارزمية الفراولة (SBA) على مجموعة البيانات التجريبية لأساليب الاختيار الثلاثة وهي (اسلوب اختيار النخبة، اسلوب اختيار عجلة الروليت، والدمج بين اسلوبي النخبة وعجلة الروليت) ومقارنتها مع الخوارزميات الشائعة،
والنتائج موضحة في الجداول (3)، (6)،(4)،(5).

الجدول (3) نتائج تطبيق خوارزمية الفراولة للأساليب الثلاثة لبيانات Slump

Slump							
Algorithm	Mother No.(droot,druu.)	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	RMSE Test	RMSE Train	RMSE Test	RMSE Train	RMSE Test
SBA	N1(0.2,2.5)	2.44888	**12.89653**	4.53622	13.33781	4.00535	**12.10960**
SBA	N1(0.4,3.5)	2.82295	**12.87265**	2.04566	**12.88434**	2.82295	**12.87265**
SBA	N1(0.6,4.5)	**1.77076**	**12.89491**	**1.89383**	**11.32979**	**1.74755**	**12.17758**
SBA	N2(0.2,2.5)	**1.95525**	**12.50862**	**1.79197**	**11.52353**	2.16203	**11.82106**
SBA	N2(0.4,3.5)	**1.89328**	**11.52131**	6.12849	**12.48207**	2.82097	**10.63014**
SBA	N2(0.6,4.5)	**1.89036**	13.48664	5.57654	**11.94798**	**1.76889**	**12.91705**
SBA	N4(0.2,2.5)	**1.82498**	13.50681	**1.77352**	**11.30080**	**1.87690**	13.18164
SBA	N4(0.4,3.5)	2.15843	12.61816	**1.72869**	**11.73197**	**1.73857**	**12.33128**
SBA	N4(0.6,4.5)	**1.71955**	**12.87262**	**1.69681**	**12.63180**	**1.84299**	**12.53445**
PSO		***1.99741***	13.84280	*1.99741*	13.84280	***1.99741***	13.84280
GA		2.84406	***13.11189***	2.84406	***13.11189***	2.84406	***13.11189***
Gride		7.29291	14.86904	7.29291	14.86907	7.29291	14.86907
SA		6.66366	13.80801	6.66366	13.80803	6.66366	13.80803

الجدول (4): نتائج تطبيق خوارزمية الفراولة للأساليب الثلاثة لبيانات Boston

Boston							
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
SBA	N1(0.2,2.5)	**3.13251**	**5.27200**	**2.60856**	**4.80969**	**2.73818**	5.00025
SBA	N1(0.4,3.5)	**3.62102**	**4.91225**	**2.63063**	**4.70784**	**3.66831**	3.78389
SBA	N1(0.6,4.5)	**2.70154**	**5.25124**	**2.87845**	**4.64666**	**2.61286**	5.62428
SBA	N2(0.2,2.5)	**2.60048**	**5.17678**	**2.63260**	**4.42864**	**3.38950**	5.96573
SBA	N2(0.4,3.5)	**2.83455**	**5.35763**	**3.04684**	**4.16946**	**2.53857**	**5.34171**
SBA	N2(0.6,4.5)	**3.74054**	5.60219	**2.73799**	**4.83329**	**3.99984**	5.78423
SBA	N4(0.2,2.5)	**4.02044**	**4.81386**	4.62884	**4.72191**	**3.84530**	4.50810
SBA	N4(0.4,3.5)	**3.43398**	**4.28857**	4.96765	**4.30002**	**3.38443**	4.44762
SBA	N4(0.6,4.5)	4.18948	**4.71065**	4.23068	**4.91484**	**3.70104**	4.64006
PSO		***4.09613***	***5.50623***	***4.09613***	***5.50623***	***4.09613***	***5.50623***
GA		4.33522	5.85656	4.33522	5.85656	4.33522	5.85656
Gride		5.46041	8.42234	5.46041	8.42234	5.46041	8.42234
SA		4.78245	6.22222	4.87542	6.22222	4.78245	6.22222

الجدول (5) نتائج تطبيق خوارزمية الفراولة للأساليب الثلاثة لبيانات Aquatic

Forestfire							
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
SBA	N1(0.2,2.5)	49.39317	**42.85173**	**19.04705**	51.75662	**28.08271**	50.65925
SBA	N1(0.4,3.5)	**33.38810**	49.85111	**29.59363**	53.36327	**35.78666**	49.68947
SBA	N1(0.6,4.5)	**38.03261**	**41.06021**	**35.63463**	**42.40550**	**28.73438**	56.49285
SBA	N2(0.2,2.5)	43.04552	48.01545	**35.76789**	54.67596	**30.12868**	46.21930
SBA	N2(0.4,3.5)	**28.37539**	49.61842	**27.07785**	46.10621	**22.32414**	**42.70503**
SBA	N2(0.6,4.5)	42.09824	**37.90466**	**28.95201**	**38.61788**	**26.79194**	47.18724
SBA	N4(0.2,2.5)	**1.82498**	**13.50683**	**2.17112**	**12.54395**	**1.78742**	**12.54069**
SBA	N4(0.4,3.5)	**2.15843**	**12.61816**	**1.97364**	**12.51096**	**1.90449**	**11.95756**
SBA	N4(0.6,4.5)	**1.71955**	**12.87262**	**1.98560**	**12.06464**	**1.70268**	**12.15999**
PSO		***40.45411***	***43.31162***	***40.45416***	***43.31162***	***40.45416***	***43.31161***
GA		43.25545	60.55880	43.25544	60.55880	43.25541	60.55880
Gride		51.69391	74.22844	48.00247	70.12445	53.85360	50.65925
SA		45.23198	57.890635	**19.04705**	51.75662	45.23198	57.89063

الجدول (6) نتائج تطبيق خوارزمية الفراولة للأساليب الثلاثة لبيانات Forestfire

						Aquatic	
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
SBA	N1(0.2,2.5)	**0.41477**	**0.85322**	0.46251	**0.93714**	**0.36493**	1.08141
SBA	N1(0.4,3.5)	**0.36759**	**0.73579**	**0.40257**	**0.86096**	**0.30902**	**0.87311**
SBA	N1(0.6,4.5)	**0.42526**	**0.86007**	0.48701	**0.89476**	**0.39953**	**0.79808**
SBA	N2(0.2,2.5)	**0.35411**	**0.93829**	**0.31281**	**0.51231**	**0.25720**	**0.68785**
SBA	N2(0.4,3.5)	**0.29497**	**0.67737**	**0.21692**	**0.54700**	**0.36822**	**0.68599**
SBA	N2(0.6,4.5)	**0.30614**	**0.98303**	**0.24029**	**0.70659**	0.50609	**0.83976**
SBA	N4(0.2,2.5)	0.57891	1.11291	0.58693	1.35948	0.53033	1.42838
SBA	N4(0.4,3.5)	**0.45273**	1.31017	0.74190	1.33992	0.52184	1.31512
SBA	N4(0.6,4.5)	0.54133	1.20878	0.53491	1.33358	0.46593	1.31881
PSO		0.48153	1.23493	0.48153	1.23493	0.48153	1.23493
GA		***0.45528***	***1.02366***	***0.45528***	***1.02366***	***0.45528***	***1.02366***

Gride		0.90467	1.68057	0.75607	1.39886	0.82895	1.55354
SA		0.70340	1.18266	0.70340	1.18266	0.70340	1.18266
							Aquatic
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
SBA	N1(0.2,2.5)	**0.41477**	**0.85322**	0.46251	**0.93714**	**0.36493**	1.08141
SBA	N1(0.4,3.5)	**0.36759**	**0.73579**	**0.40257**	**0.86096**	**0.30902**	**0.87311**
SBA	N1(0.6,4.5)	**0.42526**	**0.86007**	0.48701	**0.89476**	**0.39953**	**0.79808**
SBA	N2(0.2,2.5)	**0.35411**	**0.93829**	**0.31281**	**0.51231**	**0.25720**	**0.68785**
SBA	N2(0.4,3.5)	**0.29497**	**0.67737**	**0.21692**	**0.54700**	**0.36822**	**0.68599**
SBA	N2(0.6,4.5)	**0.30614**	**0.98303**	**0.24029**	**0.70659**	0.50609	**0.83976**
SBA	N4(0.2,2.5)	0.57891	1.11291	0.58693	1.35948	0.53033	1.42838
SBA	N4(0.4,3.5)	**0.45273**	1.31017	0.74190	1.33992	0.52184	1.31512
SBA	N4(0.6,4.5)	0.54133	1.20878	0.53491	1.33358	0.46593	1.31881
PSO		0.48153	1.23493	0.48153	1.23493	0.48153	1.23493
GA		***0.45528***	***1.02366***	***0.45528***	***1.02366***	***0.45528***	***1.02366***
Gride		0.90467	1.68057	0.75607	1.39886	0.82895	1.55354
SA		0.70340	1.18266	0.70340	1.18266	0.70340	1.18266

اظهرت النتائج التجريبية تفوق الخوارزمية المقترحة (SBA) على الخوارزميات (PSO, GA, GRID, SA) التي تمت المقارنة بها من خلال اقل جذر لمتوسط مربعات الخطأ (RMSE) كما مبينة بالجداول الاربعة اعلاه. حيث تم تمييز القيم باللون الغامق لخوارزمية الفراولة التي تمثلت بأقل RMSE، واللون الغامق وتحته خط للخوارزمية التي تمت المقارنة معها كونها تمتلك اقل RMSE مقارنة بالخوارزميات الثلاثة الشائعة الاخرى.

ومن خلال الجداول (3–6) اظهرت النتائج التجريبية تفوق الخوارزمية المقترحة (SBA) على الخوارزميات (PSO, GA, GRID, SA) التي تمت المقارنة بها من خلال اقل جذر لمتوسط مربعات الخطأ (RMSE) كما مبينة بالجداول الاربعة اعلاه. حيث تم تمييز القيم باللون الغامق لخوارزمية الفراولة التي تمثلت بأقل RMSE، واللون الغامق وتحته خط للخوارزمية التي تمت المقارنة معها كونها تمتلك اقل RMSE مقارنة بالخوارزميات الثلاثة الشائعة الاخرى.

وفضلا عن ذلك ومن خلال الجداول ((3),(4),(5),(6)) يمكن ملاحظة قيم (RMSE) لكل من بيانات التدريب والاختبار بالنسبة الى حالات خوارزمية SBA فقط، وكمقارنة فيما بينها ولجميع البيانات ولجميع أساليب الاختيار، حيث انها كانت تمتلك اقل قيم في معظم الحالات التي تكون فيها عدد العقد أكبر من واحد، وكما هو موضح في الجداول (7–10) ادناه والتي توضح الصفوف التي تحتوي على اقل قيم فقط للمعيار RMSE والتي تم تحديدها باللون الاحمر حيث ان هذه الجداول مأخوذة من الجداول (3)، (4), (5)، (6) على التوالي.

الجدول (7): تحديد اقل قيم RMSE لبيانات SLUMP

Slump							
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
SBA	N2(0.4,3.5)	**1.89328**	**11.52131**	6.12849	**12.48207**	2.82097	**10.63014**
SBA	N4(0.2,2.5)	**1.82498**	13.50681	**1.77352**	**11.30080**	**1.87690**	13.18164
SBA	N4(0.4,3.5)	2.15843	12.61816	**1.72869**	**11.73197**	**1.73857**	**12.33128**
SBA	N4(0.6,4.5)	**1.71955**	**12.87262**	**1.69681**	**12.63180**	**1.84299**	**12.53445**

الجدول (8): تحديد اقل قيم RMSE بيانات Boston

Boston							
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
SBA	N1(0.2,2.5)	**3.13251**	**5.27200**	**2.60856**	**4.80969**	**2.73818**	5.00025

SBA	N1(0.4,3.5)	**3.62102**	**4.91225**	**2.63063**	**4.70784**	**3.66831**	3.78389
SBA	N2(0.2,2.5)	**2.60048**	**5.17678**	**2.63260**	**4.42864**	**3.38950**	5.96573
SBA	N2(0.4,3.5)	**2.83455**	**5.35763**	**3.04684**	**4.16946**	**2.53857**	**5.34171**
SBA	N4(0.4,3.5)	**3.43398**	**4.28857**	4.96765	**4.30002**	**3.38443**	4.44762
SA		4.78245	6.22222	4.87542	6.22222	4.78245	6.22222

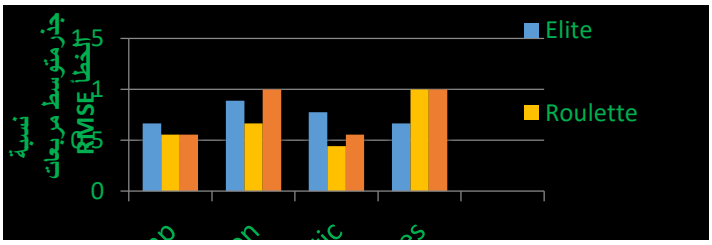
الجدول (9): تحديد اقل قيم RMSE لبيانات Aquatic

Aquatic							
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
SBA	N2(0.2,2.5)	**0.35411**	**0.93829**	**0.31281**	**0.51231**	**0.25720**	**0.68785**
SBA	N2(0.4,3.5)	**0.29497**	**0.67737**	**0.21692**	**0.54700**	**0.36822**	**0.68599**

الجدول :(10) تحديد اقل قيم RMSE لبيانات Forestfire

Forestfire							
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
SBA	N4(0.4,3.5)	**2.15843**	**12.61816**	**1.97364**	**12.51096**	**1.90449**	**11.95756**
SBA	N4(0.6,4.5)	**1.71955**	**12.87262**	**1.98560**	**12.06464**	**1.70268**	**12.15999**

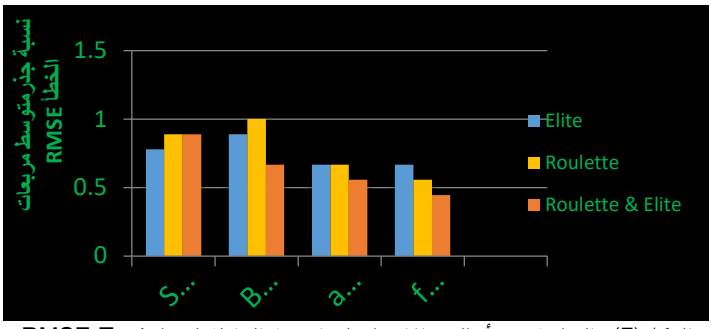
ومن اجل المقارنة بين أساليب الاختيار تم اعداد الشكل (6) والذي يوضح نتائج المقارنة بين الاساليب الثلاثة للاختيار للبيانات التجريبية الاربعة في حالة بيانات التدريب لمعيار RMSE.TRAIN ، حيث يوضح المحور الافقي الاحرف الاولى من اسماء المجاميع البيانات التجريبية.



الشكل(6): المقارنة بين أساليب الاختيار لخوارزمية الفراولة لمعيار

حيث يلاحظ من الشكل (6) أعلاه، ان هناك تفاوت في التفوق وبشكل مختلف من أسلوب اختيار الى أسلوب اختيار اخر. وان أسلوب الدمج بين اسلوبي اختيار النخبة وعجلة الروليت كان الأفضل كمعدل نسب لجميع مجاميع البيانات التجريبية بالنسبة الى المعيار RMSE.Train بامتلاكهِ اعلى نسبة وهي والتي كانت 0.78 مقارنة مع الاسلوبين الاخرين.

وبنفس الأسلوب تم اعداد الشكل (7) والذي يوضح نتائج المقارنة بين الاساليب الثلاثة للاختيار للبيانات التجريبية الاربعة في حالة بيانات الاختبار لمعيار RMSE.TEST.



الشكل(7): المقارنة بين أساليب الاختيار لخوارزمية الفراولة لمعيار RMSE.Test

يلاحظ أيضاً من الشكل (7) ان هناك تفاوت في التفوق وبشكل مختلف من أسلوب اختيار الى أسلوب اختيار اخر. وان أسلوب عجلة الروليت كان الأفضل كمعدل نسب لجميع مجاميع البيانات التجريبية بالنسبة الى المعيار RMSE.TEST بامتلاكِهِ اعلى نسبة وهي والتي كانت 0.78 مقارنة مع الاسلوبين الاخرين.

16. اختبار فريدمان The Friedman Test

وهو عبارة عن اختبار احصائي لا معلمي (Tests Nonparametric) يستخدم في حالة حجم العينة اقل من (30) مشاهدة. و اقترح استخدام هذا الاختبار في عام 1973 من قبل الباحث Friedman Milton لاختبار وجود اختلافات او فروقات معنوية في حالة وجود بيانات غير مترابطة (Porkka et al., 2008).

تم استخدام هذا الاختبار لاختبار الفروق المعنوية بين الخوارزمية المقترحة SBA مع باقي الخوارزميات الشائعة التي تم المقارنة معها للبيانات التجريبية والتطبيقية باستخدام البرنامج الاحصائي spss. والجدول (11) يوضح النتائج التي تم التوصل اليها.

الجدول (11): نتائج اختبار فريدمان لإيجاد الفروق المعنوية لخوارزمية الفراولة مع الخوارزميات الأخرى للبيانات التجريبية

Data sets	Elite Selection		Roulette selection		Roulette & Elite	
	RMSE Train	RMSE Test	RMSE Train	RMSE Test	RMSE Train	RMSE Test
Slump	Gride, SA	Non	Gride, SA	Gride, SA	Gride, SA, GA	Gride, SA
Boston	Gride, SA, GA	Gride, SA, GA	Gride, SA, GA	Gride	Gride, SA, GA	Gride
Aquatic	Gride, PSO, SA	Gride, PSO, SA	Gride, PSO, SA	Gride, PSO, SA	Gride, PSO, SA	Gride, PSO, SA
Forest fire	Gride, PSO, SA	Gride, PSO, SA	Gride, PSO, SA	Gride, PSO, SA	Gride, PSO, SA	Gride, PSO, SA

اظهرت نتائج التحليل بعد تطبيق اختبار فريدمان ومن خلال الجدول (11) ان هنالك فروقات معنوية بين الخوارزمية المقترحة (SBA) مقارنة بالخوارزميات الشائعة وهي(خوارزمية الجينية ، خوازمية سرب الطيور، خوارزمية التلدين وخوارزمية الشبكة) لبيانات التدريب وبيانات الاختبار في الاساليب الثلاثة (اسلوب اختيار النخبة ،اسلوب اختيار عجلة الروليت ، والدمج بين اسلوبي النخبة واختيار عجلة الروليت) لمجاميع البيانات التجريبية الاربعة .

17.البحث العالمي والمحلي للجانب التجريبي Global and local search in the experimental section

يقصد بالبحث العالمي(Global Search) في خوارزمية الفراولة هو البحث عن الموارد الطبيعة متمثلة بالمياه والمعادن من خلال بحث السيقان الى الاماكن الابعد. اما البحث المحلي (Local Search) هو البحث من خلال تطوير الجذور والشعيرات الجذرية الى المكان الاقرب. حيث اثبتت لنا النتائج العملية ان البحث العالمي متمثلا بالسيقان (Runners) اعطى لنا نتائج أفضل من البحث المحلي الذي يتمثل بالجذور(Roots) في الخوارزمية المقترحة. والسبب في ذلك هو ان السيقان تكون اطول من الجذور حيث يكون لها امكانية بشكل اوسع واشمل للبحث في مساحة البحث بعكس الجذور التي تكون هي الاقصر من السيقان. وقد تم اعداد جدول يحتوي على النسب التي تم ايجادها من خلال عدد مرات الحصول على RMSE للسيقان اقل مما للجذور ولكل أسلوب اختيار لكل مجموعة بيانات تجريبية وكما هو موضح في الجدول (12).

الجدول (12): نسب قيمة المعيار RMSE الاقل للسيقان مقارنة بالجذور لمجاميع البيانات التجريبية

Data Set	Elite Selection	Roulette selection	Roulette & Elite
	RMSE	RMSE	RMSE
Slump	**0.95**	**0.83**	**0.78**
Boston	**0.26**	**0.40**	**0.30**
Aquatic	**0.23**	**0.23**	**0.40**
Forestfire	**0.73**	**0.42**	**0.45**

من خلال الجدول (12) يلاحظ ان نسبة حصول السيقان على قيمة اقل لمعيار RMSE كانت نسب عالية مقارنة بالنسب التي حصلت عليها الجذور، وبشكل محدد يمكن ملاحظة ذلك من خلال مجموعة بيانات Slump و ForestFire حيث كانت تمتلك نسب مرتفعة في أساليب الاختيار وبشكل متفاوت. بينما كانت في مجموعتي بيانات Boston و Aquatic فقد كانت الجذور بشكل عام افضل في هذه البيانات، وهذا لا يعني ان السيقان لا تمثل البحث العالمي وذلك كون ان قيم الجذور كانت اصغر بشكل ملحوظ مقارنة بقيم السيقان عند اعدادها في الخوارزمية ومع هذا كانت في احد المجموعتين افضل في البحث مقارنة بالجذور.

18. الجانب التطبيقي Application part

تم في هذا الجانب استخدام بيانات حقيقة لمرضى ضغط الدم (Blood pressure) والتي تم الحصول عليها من مستشفى ازادي التعليمي في محافظة دهوك. وتتألف هذه البيانات من (17) متغير توضيحي (x) ومتغير واحد معتمد (y)، كما وتألفت العينة من (128) مشاهدة. تم اجراء التحليل عليها باستخدام خوارزمية الفراولة وباستخدام ثلاث أساليب للاختيار وهي (اسلوب اختيار النخبة، اسلوب اختيار عجلة الروليت والدمج بين اسلوبي النخبة والروليت) لاختيار أفضل توليفة من المعلمات لانحدار المتجه الداعم والنتائج موضحة بالجدول (13).

من خلال الجدول (13) ادناه اظهرت النتائج تفوق خوارزمية الفراولة على الخوارزميات الاربعة (PSO, GA, GRID, SA) . وذلك من خلال الحصول على اقل جذر لمتوسط مربعات الخطأ RMSE والتي تم تمييزها بالون الغامق لخوارزمية الفراولة، واللون الغامق وتحته خط للخوارزمية التي تمتلك اقل RMSE من الخوارزميات الثلاثة التي تمت المقارنة معها.

الجدول (13): تطبيق خوارزمية الفراولة باستخدام اساليب الاختيار الثلاثة لمرضى ضغط الدم

Algorithm	Mother No.(droot,druu.)	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	RMSE Test	RMSE Train	RMSE Test	RMSE Train	MSE Test
SBA	N1(0.2,2.5)	0.26814	**1.16645**	0.45682	**1.17510**	0.33441	**1.31164**
SBA	N1(0.4,3.5)	0.35748	**1.14983**	0.19617	**1.28512**	0.42307	**1.19895**
SBA	N1(0.6,4.5)	0.27651	**1.33886**	0.38970	**1.24616**	**0.17335**	1.38543
SBA	N2(0.2,2.5)	0.20291	**1.34731**	0.47693	**1.26526**	0.40682	**1.22637**
SBA	N2(0.4,3.5)	0.33460	**1.19411**	0.36445	**1.26472**	0.34474	**1.19996**
SBA	N2(0.6,4.5)	0.33783	**1.30731**	0.36121	**1.32029**	0.28782	**1.30056**
SBA	N4(0.2,2.5)	**0.14057**	**1.34108**	0.34093	**1.29651**	0.55718	**1.24682**
SBA	N4(0.4,3.5)	**0.13510**	1.38431	**0.11849**	**1.32253**	0.33525	**1.07293**
SBA	N4(0.6,4.5)	**0.13169**	1.42667	0.19124	**1.03522**	0.41269	**1.01593**
PSO		*0.14465*	1.36417	*0.14465*	1.36417	*0.14465*	1.36417
GA		0.14541	*1.35194*	0.14541	*1.35194*	0.14541	*1.35194*
Gride		0.69000	1.50258	0.74018	1.47638	0.77413	1.46758

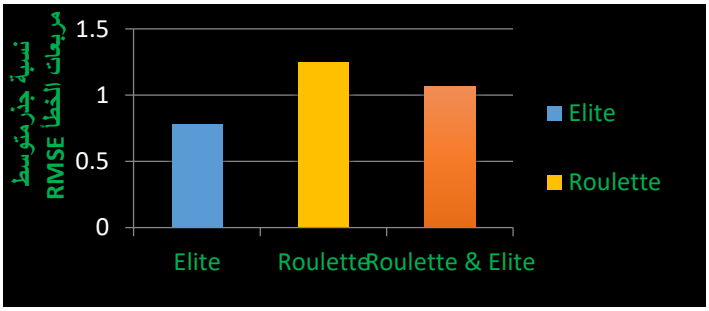
SA		0.38083	1.35437	0.38083	1.35437	0.38083	1.35437

الجدول (14) تحديد اقل قيم RMSE بيانات لمرضى ضغط الدم

ضغط الدم							
Algorithm	**Mother No.(droot,druu.)**	Elite Selection		Roulette selection		Roulette & Elite	
		RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**MSE Test**
SBA	N1(0.4,3.5)	0.35748	**1.14983**	0.19617	**1.28512**	0.42307	**1.19895**
SBA	N1(0.6,4.5)	0.27651	**1.33886**	0.38970	**1.24616**	**0.17335**	1.38543
SBA	N4(0.4,3.5)	**0.13510**	1.38431	**0.11849**	**1.32253**	0.33525	**1.07293**
SBA	N4(0.6,4.5)	**0.13169**	1.42667	0.19124	**1.03522**	0.41269	**1.01593**

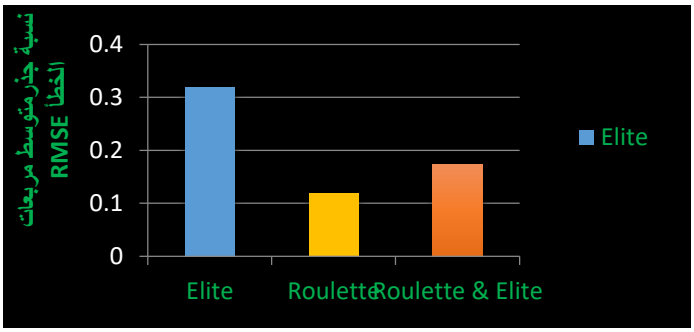
ومن اجل اثبات ان بزيادة عدد العقد لنبات الام (N) في الخوارزمية كلما كان ذلك أفضل تم اعداد الجدول (14) اعلاه. والذي من خلالهُ يلاحظ ان في حالة عدد العقد كان يساوي الواحد فان عدد الحالات التي يكون قيمة المعيار RMSE اقل هو حالتين فقط، بينما في حالة أربعة عقد فقد كان عدد القيم الذي ظهرت فيها RMSE اقل كانت أربعة قيم. بذلك يتضح من ذلك انهُ كلما زاد عدد العقد كلما كانت تعطي الخوارزمية حل أفضل وبقيمة اقل بالنسبة الى المعيار المستخدم في هذه الدراسة. وكما هو موضح في الجدول (14) والذي يوضح الصفوف التي تحتوي على اقل قيم فقط للمعيار RMSE والمؤشرة باللون الاحمر حيث ان هذا الجدول مأخوذ من الجدول (13).

ومن اجل المقارنة بين أساليب الاختيار الثلاثة لخوارزمية الفراولة لمعيار RMSE.Train لبيانات ضغط الدم تم اعداد الشكل (8) والذي يوضح تفوق اسلوب عجلة الروليت بامتلاكه اعلى نسبة وهي (1.2) مقارنة بالأسلوبين الاخرين.



الشكل (8) يوضح أساليب الاختيار لخوارزمية الفراولة لمعيار RMSE.Train لبيانات ضغط الدم

وبنفس الأسلوب للمعيار RMSE.Tes تم اعداد الشكل (9) للمقارنة بين أساليب الاختيار لخوارزمية الفراولة لمجموعة بيانات ضغط الدم، والذي يوضح تفوق اسلوب النخبة بامتلاكه اعلى نسبة وهي (0.33) مقارنة بالأسلوبين الاخرين.



الشكل (9) يوضح أساليب الاختيار لخوارزمية الفراولة لمعيار RMSE.Test لبيانات ضغط الدم

19. اختبار فريدمان The Friedman Test

تم اجراء اختبار فريدمان على البيانات الحقيقة لمرضى ضغط الدم ولذلك لاختبار الفروقات المعنوية لبيانات التدريب والاختبار كما مبينة بالجدول (15) ادناه. ومن خلال الجدول (15) اظهرت نتائج التحليل ان هنالك فروقات معنوية بين الخوارزمية المقترحة(SBA) للأساليب الثلاثة للاختيار مقارنة مع الخوارزميات الاربعة. حيث هنالك فروق معنوية بين الخوارزمية المقترحة مع خوارزمية الشبكة(Gride) في اختيار النخبة لبيانات التدريب والاختيار. وفي حالة اسلوب عجلة الروليت نلاحظ ان هنالك فروق معنوية بين خوارزمية الفراولة وخوارزميتي (Gride, SA) لبيانات التدريب والاختبار. واما في حالة الدمج بين اسلوبي اختيار النحبة واختيار عجلة الروليت لوحظ ان هنالك فروق معنوية بين الخوارزمية المقترحة والخوارزميات(Gride, PSO, SA, GA) لبيانات التدريب والاختبار.

الجدول (15): نتائج اختبار فريدمان لإيجاد الفروق المعنوية لخوارزمية الفراولة لبيانات ضغط الدم

Data sets	Elite Selection		Roulette selection		Roulette & Elite	
	RMSE Train	**RMSE Test**	**RMSE Train**	**RMSE Test**	**RMSE Train**	**RMSE Test**
ضغط الدم	Gride	Gride	Gride, SA	Gride, SA	Gride, PSO, SA, GA	Gride, PSO, SA, GA

20.البحث المحلي والعالمي في الجانب التطبيقي Global and local search in the application section

وقد تم اعداد جدول يحتوي على النسب التي تم ايجادها من خلال عدد مرات الحصول على RMSE للسيقان اقل مما للجذور ولكل أسلوب اختيار لمجموعة بيانات ضغط الدم وكما هو موضح في الجدول (16).

جدول (16): نسب قيمة المعيار RMSE الاقل للسيقان مقارنة بالجذور لمجموعة بيانات ضغط الدم

Data Set	Elite Selection	Roulette selection	Roulette & Elite
	RMSE	**RMSE**	**RMSE**
ضغط الدم	**0.78**	**0.54**	**0.69**

والجدول (16) يوضح ان السيقان أفضل من الجذور كون البحث العالمي المتمثل بالسيقان يكون أوسع في مجال المشكلة وذلك لأنها تستطيع ان تمتد الى مسافة أكثر من الجذور وبالتالي تحصل على أفضل حل مما يؤدي الى الحصول على اقل جذر متوسط مربعات خطأ اقل من الجذور المتمثلة بالبحث المحلي.

21.الاستنتاجات:

1. ان الية الربط بين خوارزمية الفراولة SBA مع نموذج انحدار المتجه الداعم SVR اعطى نتائج افضل من خلال الحصول على توليفة افضل للمعلمات الفائقة، وكان ذلك واضحاً من خلال تفوق خوارزمية الفراولة على الخوارزميات الشائعة.
2. اثبتت النتائج في الجانب التجريبي ان طريقة الدمج بين اسلوبي اختيار النخبة وعجلة الرووليت كان الأفضل. بينما كان أسلوب اختيار عجلة الروليت كان الأفضل في المعيار RMSE.Test. اما في الجانب التطبيقي فقد اثبتت النتائج ان أسلوب اختيار النخبة كان الأفضل في معيار RMSE.Train ، اما في معيار RMSE.Test فقد اثبتت النتائج ان أسلوب الاختيار عجلة الروليت فقد كان الأفضل مقارنة بالأسلوبين الاخرين.
3. عند اجراء اختبار فريدمان قد اعطى نتائج والتي يمكن الاستنتاج منها على ان خوارزمية الفراولة قد كان لها فرق معنوي مقارنة مع خورازمية Grid وSA في جميع الاختبارات تقريباً وفي معظم اساليب الاختيار ولجميع المجاميع التجريبية. واما في الجانب التطبيقي فقد أظهرت النتائج ان خوارزمية SBA كان لها فرق معنوي مع جميع الخوارزميات.
4. استخدام عدة عقد ام في خوارزمية الفراولة اعطى حل أفضل من عقدة واحدة وكان ذلك واضحا في الجانب التجريبي والتطبيقي من خلال امتلاك الحالات التي يكون عدد العقد اربعة عقد على جذر متوسط مربعات الخطأ اقل مقارنة عن باقي الحالات. وذلك لان بزيادة عدد العقد عن عقدة واحدة سيوفر مساحة بحث أكبر وبالتالي حل أفضل.

Refference

1. Kasem ,A,S.(2018)." Improving the performance of supportive vector machine classification technology using genetic algorithm", Al-Rafidain Journal of Computer Science and Mathematics,Issue 12 ,15-50.
2. AL-Kalak,A.N. & Shaaban , R.Z. (2008)." Genetic algorithm in scheduling operations with no interruption" ,AL-Rafideen Development, 89(30),245-256.
3. Aamir, N., Mushtaq, M., & Riaz, R. (2016). Effect of roots and runners in Strawberry Algorithm for optimization Problems. International Journal of Computational Analysis, 2(2), 16–21.
4. Anikesh. (2018). Types of machine learning | AnalyticsJobs. https://analyticsjobs.in/education/types-of-machine-learning.
5. Engel, Y., Mannor, S., & Meir, R. (2002). Vector Regression. Springer-Verlag Berlin Heidelberg, 84–96.
6. Evan Lutins. (2017, September 17). Grid Searching in Machine Learning: Quick Explanation and Python Implementation | by Evan Lutins | Medium. University of Maryland Graduate Current Data Scientist at Pinpoint. https://elutins.medium.com/grid-searching-in-machine-learning-quick-explanation-and-python-implementation-550552200596.
7. Jan, M. A., Asim, M., & Asad, M. (2016). Modified Strawberry Algorithms for Global Optimization Problems Modified Strawberry Algorithms for Global Optimization Problems. October.
8. Kim, J. (1997). Iterated Grid Search Algorithm on Unimodal Criteria. e University in partial fulfillmen.
9. Kumar, R., & Member, S. (2012). Blending Roulette Wheel Selection & Rank Selection in Genetic Algorithms. International Journal of Machine Learning and Computing, 2(4), 365–370.
10. Langley, P., & Carbonell, J. G. (1984). Approaches to machine learning. In Journal of the American Society for Information Science (Vol. 35, Issue 5). https://doi.org/10.1002/asi.4630350509.
11. Merrikh-Bayat, F. (2014b). A Numerical Optimization Algorithm Inspired by the Strawberry. Computer Science Neural and Evolutionary Computing, October.
12. Moriguchi, K., Ueki, T., & Saito, M. (2015). An Evaluation of the Use of Simulated Annealing to Optimize Thinning Rates for Single Even-Aged Stands. International Journal of Forestry Research, 2015(173042), 15.
13. Mu, T., & Nandi, A. K. (2009). Automatic tuning of L2-SVM parameters employing the extended Kalman filter. Expert Systems, 26(2),160–175. https://doi.org/10.1111/j.1468-0394.2009.00469.x
14. Rojas, I., Joya, G., & Catala, A. (2017). Evolutionary Support Vector Regression via Genetic Algorithms: A Dual Approach. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10306 LNCS(May), V–VII. https://doi.org/10.1007/978-3-319-59153-7.
15. Rosenbaum, L., Dörr, A., Bauer, M. R., Boeckler, F. M., & Zell, A. (2013). Inferring multi-target QSAR models with taxonomy-based multi-task learning. December. https://doi.org/10.1186/1758-2946-5-33.
16. Rosenbaum, L., Dörr, A., Bauer, M. R., Boeckler, F. M., & Zell, A. (2013). Inferring multi-target QSAR models with taxonomy-based multi-task learning. December. https://doi.org/10.1186/1758-2946-5-33.
17. Thede, S. M. (2004). An introduction to genetic algorithms *. In Consortium for Computing Sciences in Colleges. (Vol. 20, Issue 1, pp. 115–123).
18. Vapnik, ladimir N. (1995). The Nature of Statistical Learning Theory. Springer-Verlag New York.
19. Yeh, C., Huang, C., & Lee, S. (2011). Expert Systems with Applications A multiple-kernel support vector regression approach for stock market price forecasting q. Expert Systems With Applications, 38(3), 2177–2186. https://doi.org/10.1016/j.eswa.2010.08.004.

Improving machine learning prediction using strawberry algorithm

Noora Suhel Ismaeel Hawa & Salih.Mooaed Shaker
student@uomosul.edu.iq & salih.mooaed@uomosul.edu.iq
Department of Informatics & Statistic, College of Computer & Mathematical Science, University of Mosul, Mosul, Iraq
Department of Informatics & Statistic, College of Computer & Mathematical Science, University of Mosul, Mosul, Iraq.

Abstract : In this paper, the Support Vector Regression (SVR) model was used, which is defined as an algorithm or a linear model used to predict a specific model. The performance efficiency of the SVR method depends on the selection of its hyper parameters. In this paper, the SVR method was used with the Strawberry Algorithm, which is the proposed algorithm to obtain the best combination of hyper parameters.
The Root Mean Squares Error (RMSE) criterion was used to compare the results obtained using the proposed algorithm with some common algorithms, namely, Grid Search, Genetic Algorithm, Particle swarm optimization, and an annealing algorithm (Simulated Annealing algorithm. Three methods of selection were also used in the strawberry algorithm, roulette wheel selection, elite selection, and roulette wheel with the elite selection method together. The performance of the algorithm was tested through experimental and real data. The results showed that the strawberry algorithm was superior to the common algorithms in choosing the best combination of hyper parameters. The results also showed that the method of choosing the roulette wheel is the best method that gave good results compared to other methods on the experimental and applied sides.
Key Words: Support Vector Regression, Machine Learning, Algorithms, Strawberry Algorithm.